ICIC Express Letters
Volume **10**, Number **11**, November **2016**

ICIC International ©2016 ISSN 1881-803X
pp. 2607–2613

AN IMPROVED SOCIALMF MODEL BY INTEGRATING NON-SYMMETRICAL SIMILARITY INTO USERS TRUST RELATIONSHIP NETWORK

Lianhui Li[1,*] and Chunlei Mao[2]

[1]College of Mechatronic Engineering
Beifang University of Nationalities
No. 204, Wenchang North Street, Yinchuan 750021, P. R. China
*Corresponding author: llhnun@126.com

[2]Research and Development Center
Nanjing Automation Institute of Water Conservancy and Hydrology in Ministry of Water Resources
No. 95, Tiexinqiao Street, Yuhuatai District, Nanjing 210012, P. R. China
sara5778@163.com

Received March 2016; accepted June 2016

Abstract. *User's trusted users are treated equally in the recommending process by SocialMF model. To solve this problem and improve the accuracy degree of recommendation model, firstly a non-symmetrical similarity calculation method is proposed in the case of sparse score data, which can determine the existence of the same preference among trusted users, and then the preference information is integrated into the existing relationship network to strengthen the trust network. Secondly the strengthened trust network is applied to PMF model. In the decomposition process of scoring matrix, user's feature vector is affected by both trust users and the users with common preferences. Experimental results compared with the current popular models show that the proposed improved SocialMF model has better recommendation effect in the indicators RMSE and MAE.*
Keywords: Recommending process, Non-symmetrical similarity, Preference information, Trust network, User's feature vector

1. **Introduction.** With the rapid development of the Internet, the network information presents geometric growth and it is urgent to solve the problem of information overload. As an effective information filtering technology, the recommendation system has been widely concerned. Collaborative filtering, which is one of the most widely used methods of the recommendation system, is divided into memory-based and model-based methods [1-3]. To further solve the problem of data sparsity and cold start, the method of using social information among users is gradually becoming a hot research topic in the field of recommendation [4-10]. Based on PMF [4,5], it is assumed that the user's feature vector is affected by its direct trust, and SocialMF is proposed by integrating trust matrix into scoring matrix [6]. These algorithms can be modeled from the perspective of the social recommendation simulation in the real world. However, most of them treat the trust relationship between users in the process of recommendation equally. In many cases, there is not a common interest between the two users who have a trust relationship. Therefore, a new method to measure the similarity of users is presented in this research, and applied in SocialMF model to solve this problem. The remainder of this paper is organized as follows. Section 2 proposes the improved SocialMF model based on the non-symmetrical similarity measuring method. Implementation of the improved SocialMF model is described in Section 3. The recommendation performance simulation experiments of the proposed improved SocialMF model, UserItemBias, PMF, AC-PMF and SocialMF algorithms are given in Section 4. Finally, conclusions are presented in Section 5.

2. **Improved SocialMF Model.**

2.1. **Non-symmetrical similarity measuring method.**

In sparse matrix, due to the less common scoring items of the user, the error of the similarity obtained by traditional similarity calculation method is larger. Aimed at the above shortcomings, a non-symmetrical similarity measuring method is proposed. The similarity between the two users is not considered from the overall point of view. With the comprehensive consideration of the user's score of all items, the non-symmetrical coefficient formula is designed as: when $|Y_i| \geq \bar{Y}$, $NSC(i,j) = |Y_i \cap Y_j|/|Y_i|$; when $|Y_i| < \bar{Y}$, $NSC(i,j) = |Y_i \cap Y_j|/(\bar{Y} - |Y_i|)$. Here, Y_i represents the set of scoring items of the user U_i, $|Y_i|$ represents the number of scoring items of the user U_i, and $\bar{Y}$ represents the number of common scoring items of all users. Therefore, the non-symmetrical similarity measuring formula is obtained as $S_{ij} = NSC(i,j) \cdot SIM(i,j)$, where:

$$SIM(i,j) = \sum_{k \in Y_{ij}} (R_{ik} - \bar{R}_i)(R_{jk} - \bar{R}_j) \Bigg/ \sqrt{\sum_{k \in Y_{ij}} (R_{ik} - \bar{R}_i)^2} \sqrt{\sum_{k \in Y_{ij}} (R_{jk} - \bar{R}_j)^2} \tag{1}$$

Here, $\bar{R}_i$ represents the average score of all items by the user U_i, Y_{ij} represents the set of common scoring items of the users U_i and U_j, and $\bar{Y}$ represents the number of common scoring items of all users, $S_{ij} \in [-1, 1]$.

2.2. **The recommendation model strengthening the user trust relationship.**

In SocialMF, only the trust relationship between users is considered [11]. Table 1 shows an example of trust relationship and non-symmetrical similarity between two users. We get the following based on SocialMF [6].

$$\begin{bmatrix} \tilde{U}_1 \\ \tilde{U}_2 \\ \tilde{U}_3 \\ \tilde{U}_4 \\ \tilde{U}_5 \end{bmatrix} = \begin{bmatrix} 0 & 1/3 & 1/3 & 1/3 & 0 \\ 1/2 & 0 & 0 & 0 & 1/2 \\ 1/2 & 0 & 0 & 1/2 & 0 \\ 1/3 & 0 & 1/3 & 0 & 1/3 \\ 1/4 & 1/4 & 1/4 & 1/4 & 0 \end{bmatrix} \begin{bmatrix} U_1 \\ U_2 \\ U_3 \\ U_4 \\ U_5 \end{bmatrix} \tag{2}$$

In Formula (2), each user's trusted users are treated equally in the recommending process by SocialMF [6], which has a deviation from the actuality. Based on the above analysis, the non-symmetric similarity information among the trusted users is integrated into the trust network, and the user's feature vector can be obtained as $\tilde{U}_u = \sum_{v \in N_u} (T_{uv} + S_{uv}) \cdot U_v$, where $\tilde{U}_u$ is the feature vector of the user u being estimated, S_{uv} is the non-symmetrical similarity of the user u to the user v. S_{uv} only exists between two trusted users. The corresponding probability graph model is shown in Figure 1.

Table 1. The trust relationship and non-symmetrical similarity between two users

User	*His trusted user*				
	U_1	U_2	U_3	U_4	U_5
U_1	/	0.2	-0.1	-0.25	Null
U_2	0.125	/	Null	Null	0.25
U_3	-0.2	Null	/	0.3	Null
U_4	-0.333	Null	0.1	/	0
U_5	0.5	0.125	0.6	0	/

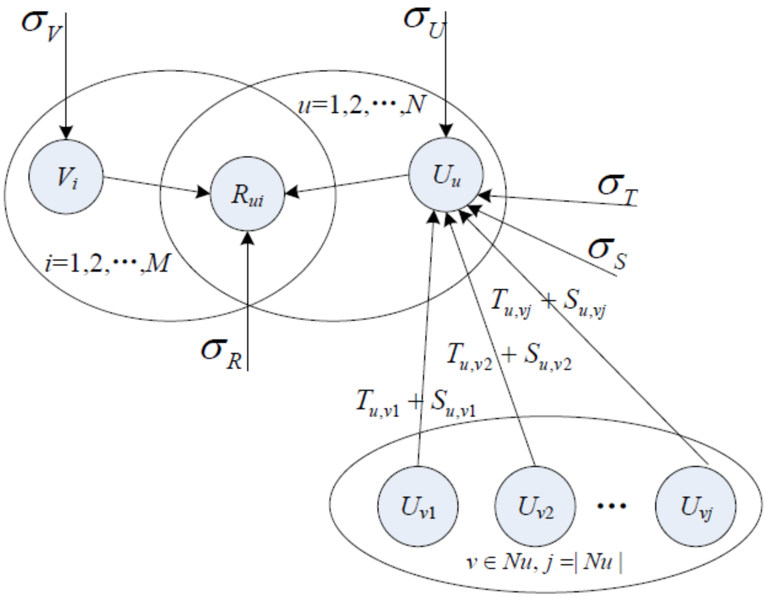

Figure 1. Probability graph model

It can be obtained combined with the information in Table 1:

$$\begin{bmatrix} \tilde{U}_1 \\ \tilde{U}_2 \\ \tilde{U}_3 \\ \tilde{U}_4 \\ \tilde{U}_5 \end{bmatrix} = \begin{bmatrix} 0 & 15/8 & 30/7 & 1/12 & 0 \\ 5/8 & 0 & 0 & 0 & 3/4 \\ 3/10 & 0 & 0 & 4/5 & 0 \\ 0 & 0 & 13/30 & 0 & 1/3 \\ 3/4 & 3/8 & 17/20 & 1/4 & 0 \end{bmatrix} \begin{bmatrix} U_1 \\ U_2 \\ U_3 \\ U_4 \\ U_5 \end{bmatrix} \tag{3}$$

3. Implementation of Improved SocialMF Model.

3.1. Conditional distribution of score matrix. The conditional probability distribution of R is defined as $p\left(R|U,V,\sigma_R^2\right) = \prod_{u=1}^{N}\prod_{i=1}^{M}\left(N\left(R_{u,i}|g\left(U_u^TV_i\right),\sigma_R^2\right)\right)^{I_{u,i}^R}$.

3.2. Conditional distribution of user feature vector. Assuming that the non-symmetrical similarity is independent of the trust degree among the trust users and obeys Gauss distribution, it is obtained as: $p(U|T,S,\sigma_U^2,\sigma_T^2,\sigma_S^2) \propto p(U|\sigma_U^2)\times p(U|T,\sigma_T^2)\times p(U|S,\sigma_S^2) = \prod_{u=1}^{N} N(U_u|0,\sigma_U^2\mathrm{I}) \times \prod_{u=1}^{N} N\left(U_u\middle|\sum_{v\in N_u} T_{u,v}U_v,\sigma_T^2\mathrm{I}\right) \times \prod_{u=1}^{N} N\left(U_u\middle|\sum_{v\in N_u} S_{u,v}U_v,\sigma_S^2\mathrm{I}\right)$.

3.3. Joint probability distribution of user feature matrix and item feature matrix. Assuming that the item feature matrix V obeys Gauss distribution with a mean value of zero, it is obtained according to Bayes theory:

$$p(V|\sigma_V^2) = \prod_{i=1}^{M} N(V_i|0,\sigma_R^2\mathrm{I}) \tag{4}$$

$$\begin{aligned} &p(U,V|R,T,S,\sigma_R^2,\sigma_U^2,\sigma_V^2,\sigma_T^2,\sigma_S^2) \propto p(R|U,V,\sigma_R^2)\times p(U|T,S,\sigma_U^2,\sigma_T^2,\sigma_S^2)\times p(V|\sigma_V^2) \\ &= \prod_{u=1}^{N}\prod_{i=1}^{M}(N(R_{u,i}|g(U_u^TV_i),\sigma_R^2))^{I_{u,i}^R} \times \prod_{u=1}^{N} N(U_u|0,\sigma_U^2\mathrm{I}) \times \prod_{u=1}^{N} N\left(U_u\middle|\sum_{v\in N_u} T_{u,v}U_v,\sigma_T^2\mathrm{I}\right) \\ &\quad \times \prod_{u=1}^{N} N\left(U_u\middle|\sum_{v\in N_u} S_{u,v}U_v,\sigma_S^2\mathrm{I}\right) \times \prod_{u=1}^{N} N(V_i|0,\sigma_V^2\mathrm{I}) \end{aligned} \tag{5}$$

The logarithm joint posterior probability distribution of the parameter U, V is:

$$\begin{aligned}
&\ln p\left(U, V | R, T, S, \sigma_R^2, \sigma_U^2, \sigma_V^2, \sigma_T^2, \sigma_S^2\right) \\
=& -\frac{1}{2\sigma_R^2}\sum_{u=1}^{N}\sum_{i=1}^{M} I_{u,i}^R\left(R_{u,i} - g\left(U_u^T V_i\right)\right)^2 - \frac{1}{2\sigma_U^2}\sum_{u=1}^{N} U_u^T U_u - \frac{1}{2\sigma_V^2}\sum_{i=1}^{M} V_i^T V_i \\
& -\frac{1}{2\sigma_T^2}\sum_{u=1}^{N}\left(U_u - \sum_{v\in N_u} T_{u,v}U_v\right)^T\left(U_u - \sum_{v\in N_u} T_{u,v}U_v\right) \\
& -\frac{1}{2\sigma_S^2}\sum_{u=1}^{N}\left(U_u - \sum_{v\in N_u} S_{u,v}U_v\right)^T\left(U_u - \sum_{v\in N_u} S_{u,v}U_v\right) - \frac{1}{2}\left(\sum_{u=1}^{N}\sum_{i=1}^{M} I_{u,i}^R\right)\ln\sigma_R^2 \\
& -\frac{1}{2}\left(N\cdot K\cdot\left(\ln\sigma_U^2 + \ln\sigma_T^2 + \ln\sigma_S^2\right) + M\cdot K\cdot\ln\sigma_V^2\right) + C
\end{aligned} \tag{6}$$

where C is a constant that has no relationship with the parameters. To maximize the posterior probability when the parameters are fixed is equivalent to minimize the error square sum function of the following with regular terms:

$$\begin{aligned}
& L(R, U, V, T, S) \\
=& \frac{1}{2}\sum_{u=1}^{N}\sum_{i=1}^{M} I_{u,i}^R\left(R_{u,i} - g(U_u^T V_i)\right)^2 + \frac{\lambda_U}{2}\sum_{u=1}^{N} U_u^T U_u + \frac{\lambda_V}{2}\sum_{i=1}^{M} V_i^T V_i \\
& +\frac{\lambda_T}{2}\sum_{u=1}^{N}\left(U_u - \sum_{v\in N_u} T_{u,v}U_v\right)^T\left(U_u - \sum_{v\in N_u} T_{u,v}U_v\right) \\
& +\frac{\lambda_S}{2}\sum_{u=1}^{N}\left(U_u - \sum_{v\in N_u} S_{u,v}U_v\right)^T\left(U_u - \sum_{v\in N_u} S_{u,v}U_v\right)
\end{aligned} \tag{7}$$

where $\lambda_U = \sigma_R^2/\sigma_U^2$, $\lambda_V = \sigma_R^2/\sigma_V^2$, $\lambda_T = \sigma_R^2/\sigma_T^2$, and $\lambda_S = \sigma_R^2/\sigma_S^2$. Local minimum value can be obtained by using the gradient descent method to solve the objective function:

$$\frac{\partial L}{\partial V_i} = \sum_{u=1}^{N} I_{u,i}^R U_v g'\left(U_u^T V_i\right)\left(g\left(U_u^T V_i\right) - R_{u,i}\right) + \lambda_V V_i \tag{8}$$

$$\begin{aligned}
\frac{\partial L}{\partial U_u} =& \sum_{i=1}^{M} I_{u,i}^R V_i g'\left(U_u^T V_i\right)\left(g\left(U_u^T V_i\right) - R_{u,i}\right) + \lambda_U U_u + \lambda_T\left(U_u - \sum_{v\in N_u} T_{u,v}U_v\right) \\
& -\lambda_T\sum_{\{v|u\in N_v\}} T_{v,u}\left(U_v - \sum_{w\in N_v} T_{v,w}U_w\right) + \lambda_S\left(U_u - \sum_{v\in N_u} S_{u,v}U_v\right) \\
& -\lambda_S\sum_{\{v|u\in N_v\}} S_{v,u}\left(U_v - \sum_{w\in N_v} S_{v,w}U_w\right)
\end{aligned} \tag{9}$$

where $g'(x) = e^{-x}/(1+e^{-x})^2$ is the first-order derivative of the logic function $g(x) = 1/(1+e^{-x})$.

3.4. **Algorithm description of the proposed improved SocialMF model.** The input includes the score matrix $R \in R^{N\times M}$, the trust matrix $T \in T^{N\times N}$, the vector feature dimension k, the learning efficiency η and the penalty parameter λ_U, λ_V, λ_T, λ_S and the outputs are the feature vector matrices of user and item. The feature matrices of user and item are randomly generated obeying normal distribution. The detailed algorithm description of the proposed improved SocialMF model is shown in Figure 2.

4. **Simulation Experiment.** Through experiments, the recommendation performance of the proposed improved SocialMF model, UserItemBias [8], PMF [9], AC-PMF [10] and SocialMF [11] are compared, and the effect of the experimental parameter λ_S on the recommended result in the proposed improved SocialMF model is analyzed.

while error on validation set decrease **do**
 for each $u \in U$ **do**
  $x = U_u - \sum_{c \in N_u} T_{u,v} U_v;$
  $y = \sum_{v|u \in N_v} T_{u,v}(U_v - \sum_{w \in N_v} T_{v,w} U_w);$
  $\tilde{x} = U_u - \sum_{c \in N_u} S_{u,v} U_v;$
  $\tilde{y} = \sum_{v|u \in N_v} S_{u,v}(U_v - \sum_{w \in N_v} S_{v,w} U_w);$
  for each $i \in V|u,i \in R$ **do**
   $\tilde{r}_{ui} = U_u^T V_i;$
   $e = r_{ui} - \tilde{r}_{ui};$
   for w=1, 2, …, $|N_v|$ **do**
    $U_{uw} -= \eta \cdot (e \cdot V_{iw} + \lambda_U \cdot U_{uw} + \lambda_T \cdot x_w - \lambda_T \cdot y_w + \lambda_S \cdot x_w - \lambda_S \cdot y_w);$
    $V_{iw} -= \eta \cdot (e \cdot U_{uw} + \lambda_V \cdot V_{iw});$
   end for
  end for
 end for
end while

Figure 2. The detailed algorithm description

4.1. **Data set and evaluation criteria.** Epinions data set contains 664824 scores of 139738 items by 49290 users and 487181 trust relationships. To evaluate these algorithms, this research uses two kinds of common indicators: root mean square error (RMSE) and mean absolute error (MAE). The formula of RMSE is:

$$RMSE = \sqrt{\sum_{(u,i)|R_{test}} (r_{u,i} - \tilde{r}_{u,i})^2 \Big/ R_{test}} \tag{10}$$

where R_{test} indicates the number of test samples, and $r_{u,i}$ and $\tilde{r}_{u,i}$ indicate the true value and predictive value of the score of the item i by the user u, respectively. Similar to the definition of RMSE, the formula of MAE is:

$$MAE = \sum_{(u,i)|R_{test}} |r_{u,i} - \tilde{r}_{u,i}| \Big/ R_{test} \tag{11}$$

4.2. **Result comparison.** Experiments are carried out on Epinions data sets which are randomly selected and the ratios of training and test sets are 8/2, 9/1. The parameters used in the experiment are as follows: $\lambda_S = 9$, $\lambda_T = 1$, $\lambda_U = \lambda_V = 0.1$. When the dimensions of feature vectors are five and ten respectively, the experimental results are shown in Table 2.

As shown in Table 2, it can be seen that the model-based recommendation algorithm is significantly better than the memory-based recommendation algorithm in the case of large amount and sparse data. The non-symmetrical similarity strengthens the trust network and improves the shortage of SocialMF.

The influence of the parameter λ_S on MAE and RMSE when the dimension of feature vector is five is shown in Figure 3. The changing situation of the recommendation result with λ_S reflects the effect of the improved trust network on the recommendation result.

4.3. **The practicability of non-symmetrical similarity in probability matrix decomposition model.** To verify the applicability of non-symmetrical similarity in SocialMF, Pearson similarity and non-symmetrical similarity are integrated into the trust

TABLE 2. The comparison of experimental results

Dimension		5		10	
Training set ratio		90%	80%	90%	80%
UserItemBias	MAE	0.8175	0.8249	0.8175	0.8233
	RMSE	1.0510	1.0593	1.0511	1.0571
PMF	MAE	0.8249	0.8313	0.8295	0.8307
	RMSE	1.0498	1.0537	1.0499	1.0510
SocialMF	MAE	0.8125	0.8156	0.8129	0.8159
	RMSE	1.0417	1.0462	1.0415	1.0461
AC-PMF	MAE	0.8125	0.8168	0.8126	0.8167
	RMSE	1.0393	1.0448	1.0394	1.0443
Improved SocialMF	MAE	0.8105	0.8140	0.8108	0.8126
	RMSE	1.0383	1.0437	1.0386	1.0435

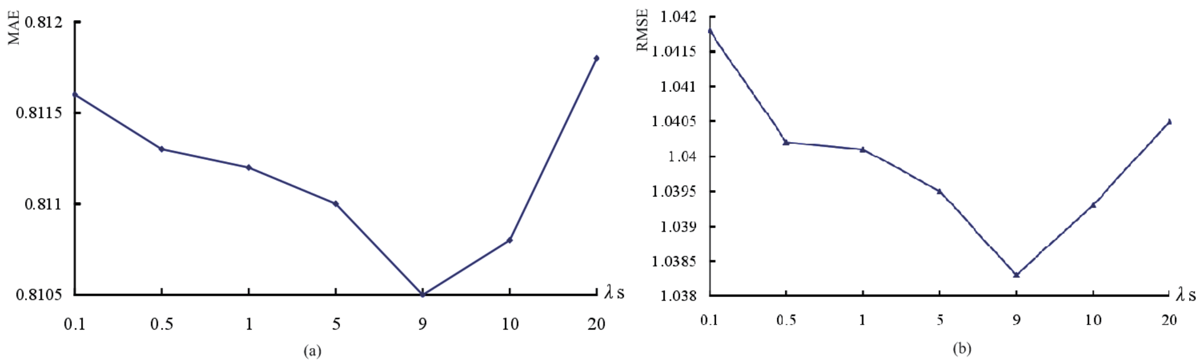

FIGURE 3. The influence of the parameter λ_S on MAE and RMSE

TABLE 3. The comparison results

Dimension	Training set ratio	Indicator	Pearson similarity	Non-symmetrical similarity
5	90%	MAE	0.8900	0.8105
		RMSE	1.1570	1.0393
	80%	MAE	0.8900	0.8140
		RMSE	1.1580	1.0437
10	90%	MAE	0.8900	0.8108
		RMSE	1.1570	1.0386
	80%	MAE	0.8900	0.8126
		RMSE	1.1580	1.0435

network, respectively. The comparison experiments are shown in Table 3 based on ScoialMF.

With the integration of Pearson similarity, overfitting phenomenon appears after the first iteration under different conditions, which shows that Pearson similarity is not suitable. However, with the integration of non-symmetrical similarity, MAE and RMSE tend to be stable after reaching the optimal value, which proves that non-symmetrical similarity is applicable.

5. **Conclusions.** By strengthening the relationship among trust users, those who have similar interests are more likely to recommend items to each other. As shown in the experiment results on Epinions data set, the integration of non-symmetrical similarity into the trust network is conducive to strengthening the relationship among trusted users

which plays an important role in the recommendation process. How to use the context information (user comment, time, class, etc.) to further improve the recommendation effect will be studied next.

Acknowledgements. This work is supported by Science Research Project of Ningxia High Education (No. NGY2016141).

REFERENCES

[1] D. Mukund and G. Karypis, Item-based top-N recommendation algorithms, *ACM Trans. Information Systems*, vol.22, no.1, pp.143-177, 2004.
[2] B. Sarwar, G. Karypis, J. Konstan and J. Riedl, Item-based collaborative filtering recommendation algorithms, *Proc. of the 10th International Conference on World Wide Web*, pp.285-295, 2001.
[3] R. Salakhutdinov and A. Mnih, Probabilistic matrix factorization, *Proc. of Advances in Neural Information Processing Systems*, pp.1257-1264, 2007.
[4] H. Ma, H. Yang, M. R. Lyu and I. King, SoRec: Social recommendation using probabilistic matrix factorization, *Proc. of the 17th ACM Conference on Information and Knowledge Management*, Napa Valley, CA, USA, pp.931-940, 2008.
[5] H. Ma, I. King and M. R. Lyu, Learning to recommend with social trust ensemble, *Proc. of the 32nd International ACM SIGIR conference on Research and Development in Information Retrieval*, pp.203-210, 2009.
[6] M. Jamali and M. Ester, A matrix factorization technique with trust propagation for recommendation in social networks, *Proc. of the 2010 ACM Conference on Recommender Systems*, Barcelona, Spain, pp.1055-1066, 2010.
[7] L. Guo, J. Ma and Z. Chen, Trust strength aware social recommendation method, *Journal of Computer Research & Development*, vol.50, no.9, pp.1805-1813, 2013.
[8] C. Chen, J. Zeng, X. Zheng and D. Chen, Recommender system based on social trust relationships, *Proc. of the 2013 IEEE 10th International Conference on E-Business Engineering*, pp.32-37, 2013.
[9] H. Fang, Y. Bao and J. Zhang, Leveraging decomposed trust in probabilistic matrix factorization for effective recommendation, *Proc. of the 28th AAAI Conference on Artificial Intelligence*, pp.30-36, 2014.
[10] K. Chen, Q. Zou, Z. P. Peng and W. D. Ke, Collaborative ranking friend recommendation algorithm in heterogeneous social network, *Journal of Chinese Computer Systems*, vol.35, no.6, pp.1270-1274, 2014.
[11] A. Anderson, D. Huttenlocher, J. Kleinberg and J. Leskovec, Effects of user similarity in social media, *Proc. of the 5th International Conference on Web Search and Web Data Mining*, Seattle, WA, USA, pp.703-712, 2012.